

Expectations of hook products on large partitions and the chi-square distribution

Mark Adler*, Alexei Borodin, Pierre van Moerbeke†

(Communicated by Peter Sarnak)

Abstract. Given uniform probability on words of length $M = Np + k$, from an alphabet of size p , consider the probability that a word (i) contains a subsequence of letters $p, p - 1, \dots, 1$ in that order and (ii) that the maximal length of the disjoint union of $p - 1$ increasing subsequences of the word is $\leq M - N$. A generating function for this probability has the form of an integral over the Grassmannian of p -planes in \mathbb{C}^n . The present paper shows that the asymptotics of this probability, when $N \rightarrow \infty$, is related to the k^{th} moment of the χ^2 -distribution of parameter $2p^2$. This is related to the behavior of the integral over the Grassmannian $\text{Gr}(p, \mathbb{C}^n)$ of p -planes in \mathbb{C}^n , when the dimension of the ambient space \mathbb{C}^n becomes very large. A different scaling limit for the Poissonized probability is related to a new matrix integral, itself a solution of the Painlevé IV equation. This is part of a more general set-up related to the Painlevé V equation.

2000 Mathematics Subject Classification: 60C55, 60F05, 05A05, 05A15, 34M55.

Contents

1 Introduction	160
2 A differential equation for a matrix integral	166
3 Hermitian matrix integrals and Painlevé equations	171
4 A limit theorem for the probability on words and Chi-square	176
5 A limit theorem for the Poissonized probability on words	178
6 Painlevé IV as a limit of Painlevé V	182
7 Appendix: Chazy classes	184

* The support of a National Science Foundation grant # DMS-01-00782 is gratefully acknowledged.

† The support of a National Science Foundation grant # DMS-01-00782, a Nato, a FNRS and a Francqui Foundation grant is gratefully acknowledged. This work was done while PvM was a member of the Clay Mathematics Institute, One Bow Street, Cambridge, MA 02138, USA. MA and PvM would like to thank Persi Diaconis for interesting conversations regarding section 4.

1 Introduction

Consider the set of words

$$\pi \in S_\ell^p := \{\text{words } \pi \text{ of length } \ell, \text{ built from an alphabet } \{1, \dots, p\}\},$$

(with $|S_\ell^p| = p^\ell$) with the uniform probability distribution

$$(1.0.1) \quad P^{\ell,p}(\pi) = \frac{1}{p^\ell}.$$

Let \mathbb{Y} denote the set of all partitions λ . Let λ^\top be the dual partition, i.e., obtained by flipping the Young diagram λ about its diagonal. So, λ_1^\top is the length of the first column of λ . Moreover h^λ denotes the product of the hook lengths $h_{ij}^\lambda := \lambda_i + \lambda_j^\top - i - j + 1$ and also 1^p denotes the infinite vector

$$1^p := (\overbrace{1, \dots, 1}^p, 0, 0, \dots).$$

The RSK correspondence between words and pairs of semi-standard and standard tableaux induces a probability measure on partitions

$$(1.0.2) \quad \lambda \in \mathbb{Y}_\ell = \{\text{partitions } \lambda \in \mathbb{Y} \text{ of weight } |\lambda| = \ell\},$$

given by

$$(1.0.3) \quad P^{\ell,p}(\lambda) = \frac{f^\lambda \mathbf{s}_\lambda(1^p)}{p^{|\lambda|}},$$

having

$$(\text{support } P^{\ell,p}) \subseteq \mathbb{Y}_\ell^{(p)} := \{\lambda \in \mathbb{Y}_\ell, \text{ such that } \lambda_1^\top \leq p\}.$$

The symbol \mathbf{s}_λ denotes the Schur polynomial associated with the partition λ . Besides the probability $P^{\ell,p}$, we also consider the corresponding Poissonized measure, depending on the real parameter x ,

$$(1.0.4) \quad P_{x,p}(\lambda) = e^{-px} \frac{(xp)^{|\lambda|}}{|\lambda|!} P^{|\lambda|,p}(\lambda), \quad \lambda \in \mathbb{Y}^{(p)},$$

having

$$(\text{support } P_{x,p}) \subseteq \mathbb{Y}^{(p)} := \{\lambda \in \mathbb{Y}, \text{ such that } \lambda_1^\top \leq p\}.$$

We discuss very briefly the combinatorics needed in this problem; for more details, see [17, 19, 20]. For the partition λ , define the symbol

$$(1.0.5) \quad (n)_\lambda := \prod_i (n+1-i)_{\lambda_i}, \quad \text{with } (x)_n = x(x+1)\dots(x+n-1), \quad x_0 = 1$$

Define for $q \geq \lambda_1^\top$, (throughout Δ_q denotes the Vandermonde determinant in q variables)

$$(1.0.6) \quad f^\lambda = \#\{\text{standard tableaux of shape } \lambda\} \\ = \frac{|\lambda|!}{h^\lambda} = \frac{|\lambda|!}{u^{|\lambda|}} \mathbf{s}_\lambda(x) |_{\sum_i x_i^i = \delta_{1i} u} = |\lambda|! \frac{\Delta_q(q + \lambda_1 - 1, \dots, q + \lambda_q - q)}{\prod_{i=1}^q (q + \lambda_i - i)!},$$

and

$$(1.0.7) \quad \#\left\{ \begin{array}{l} \text{semi-standard tableaux} \\ \text{of shape } \lambda \text{ filled with} \\ \text{numbers from 1 to } q \end{array} \right\} = \mathbf{s}_\lambda(1^q) \\ = \frac{\Delta_q(q + \lambda_1 - 1, \dots, q + \lambda_q - q)}{\prod_{i=1}^{q-1} i!} = \prod_{(i,j) \in \lambda} \frac{j - i + q}{h_{ij}^\lambda} = \frac{(q)_\lambda}{h^\lambda}$$

A subsequence σ of the word π is *weakly k -increasing*, if it can be written as

$$(1.0.8) \quad \sigma = \sigma_1 \cup \sigma_2 \cup \dots \cup \sigma_k,$$

where σ_i are disjoint weakly increasing subsequences of the word π , i.e., possibly with repetitions. The length of the longest increasing/decreasing subsequences is closely related to the shape of the associated partition, via the RSK correspondence:

$$(1.0.9) \quad d_1(\pi) = \left\{ \begin{array}{l} \text{length of the longest strictly} \\ \text{decreasing subsequence of } \pi \end{array} \right\} = \lambda_1^\top \\ i_k(\pi) = \left\{ \begin{array}{l} \text{length of the longest weakly} \\ k\text{-increasing subsequence of } \pi \end{array} \right\} = \lambda_1 + \dots + \lambda_k$$

Define the generalized hypergeometric function in terms of the symbol (1.0.5), viewed as a symmetric function in an infinite number of variables x_i

$$(1.0.10) \quad {}_2F_1^{(1)}(p, q; n; x) := \sum_{\kappa \in \mathbb{Y}} \frac{(p)_\kappa (q)_\kappa}{(n)_\kappa} \frac{\mathbf{s}_\kappa(x)}{h^\kappa},$$

which, upon restriction, using power sums and upon using (1.0.6), yields

$$\begin{aligned}
 (1.0.11) \quad {}_2F_1^{(1)}(p, q; n; x) \Big|_{\sum_i x_i^j = \delta_{1j} u} \\
 = \sum_{\kappa \in \mathbb{Y}} u^{|\kappa|} \frac{(p)_\kappa (q)_\kappa}{(h^\kappa)^2 (n)_\kappa} = \sum_{k=0}^{\infty} u^k \sum_{\kappa \in \mathbb{Y}_k} \frac{(p)_\kappa (q)_\kappa}{(h^\kappa)^2 (n)_\kappa}
 \end{aligned}$$

As a reminder, the chi-square distribution of parameter m is the distribution of

$$Z_m = \sum_{i=1}^m X_i^2,$$

where the X_i 's are m independent normal $N(0, 1)$ -random variables.

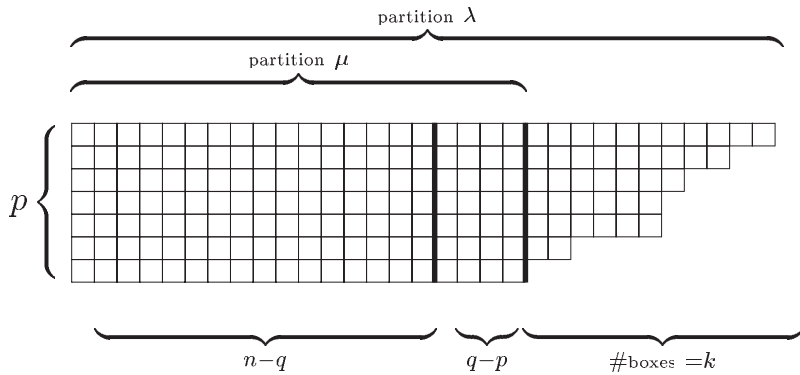


Fig. 1

The expectations $E_{x,p}$ and $E^{\ell,p}$ are taken with regard to the probabilities $P_{x,p}$ and $P^{\ell,p}$ defined above. The functions on partitions, of which the expectations are taken, are products of hook lengths restricted to a vertical strip in the partition of width $q-p$, as in Figure 1. In [3], expectations of this type have been studied and linked to integrals over the Grassmannian space $\text{Gr}(p, \mathbb{C}^n)$ of p -planes in \mathbb{C}^n , with regard to the Weyl measure $d\rho(Z)$; besides, these integrals relate to specific solutions of the Painlevé V equation.

The motivation came from the fact that expanding (with regard to a parameter) certain integrals over the unitary group $U(n)$ led to interesting combinatorics related to longest increasing sequences of random permutations and words, and also to percolation problems; see [22, 2]. During the last ten years, many striking results have been obtained on the asymptotics of the length of the longest increasing sequences in random words and random permutations. More generally, what happens with integrals over symmetric spaces, rather than groups. The simplest symmetric spaces are the Grassmannian spaces $\text{Gr}(p, \mathbb{C}^n)$. In this case, the integrals considered are Fourier transforms of the sample canonical correlation distribution for two Gaussian pop-

ulations in multivariate statistics; this distribution is known to be related to the geometry of planes in \mathbb{C}^n ; see [18, 3].

We now state the following proposition, established in [3]:

Proposition 1.1. *Given the partition*

$$(1.0.12) \quad \mu = (n-p)^p := (\overbrace{n-p, n-p, \dots, n-p}^p)$$

the mathematical expectation of the product of the hook lengths over the strip of width $q-p$, with regard to the probabilities (1.0.4) and (1.0.3), is given by¹

$$\begin{aligned} (1.0.13) \quad e^{px} E_{x,p} & \left(I_{\{\lambda \supseteq \mu\}}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) \\ &= \sum_{\ell \geq p(n-p)} \frac{(px)^\ell}{\ell!} E^{\ell,p} \left(I_{\{\lambda \supseteq \mu\}}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) \\ &= \tilde{\mathbf{c}} x^{(n-p)p} {}_2F_1^{(1)}(p, q; n; y) \big|_{\sum_\ell y_\ell^i = \delta_{1i} x} \\ &= \tilde{\mathbf{c}} x^{(n-p)p} \int_{\text{Gr}(p, \mathbb{C}^n)} e^{x \text{Tr}(I + Z^\dagger Z)^{-1}} \det(Z^\dagger Z)^{-(q-p)} d\rho(Z) \\ &= \tilde{\mathbf{c}} x^{(n-p)p} \exp \int_0^x \frac{u(y) - p(n-p) + py}{y} dy, \end{aligned}$$

where² $u(x)$ is the unique solution to the initial value problem:

$$(1.0.14) \quad \begin{cases} x^2 u''' + xu'' + 6xu'^2 - 4uu' + 4Qu' - 2Q'u + 2R = 0 \\ \text{with} \\ u(x) = p(n-p) - \frac{p(n-q)}{n}x + \dots + a_{n+1}x^{n+1} + O(x^{n+2}), \quad \text{near } x = 0, \end{cases} \quad (\text{Painlevé V})$$

with a specific coefficient a_{n+1} and where

¹ Remember (support $P^{\ell,p}$) $\subseteq \{\lambda \in \mathbb{Y}, \text{ such that } |\lambda| = \ell, \lambda_1^\top \leq p\}$.

² With

$$\tilde{\mathbf{c}}^{-1} := \prod_{i=1}^p i!(n-q-i)! \quad \text{and} \quad \tilde{\mathbf{c}} := \frac{(q-i)!}{(n-i)!}.$$

$$(1.0.15) \quad 4Q = -x^2 + 2(n + 2(p - q))x - (n - 2p)^2$$

$$2R = p(p - q)(x + n - 2p).$$

This paper is concerned with what happens when $n \rightarrow \infty$. This is related to the behavior of the integral over the Grassmannian $\text{Gr}(p, \mathbb{C}^n)$ of p -planes in \mathbb{C}^n , when the dimension of the ambient space \mathbb{C}^n becomes very large. To be precise:

Theorem 1.2. *Given the partition $\mu = (n - p)^p$, as in Figure 1, the following expectation behaves, for large n , like the moments of the chi-square distribution of parameter $2pq$:*

$$(1.0.16) \quad \lim_{n \rightarrow \infty} n^{(p^2-1)/2} E^{p(n-p)+k, p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{n-q < j \leq n-p} h_{ij}^\lambda \right)$$

$$= \frac{\sqrt{p} \prod_1^p (q - j)!}{(\sqrt{2\pi})^{p-1} k!} E \left(\left(\frac{1}{2} Z_{2pq} \right)^k \right),$$

and so the expectation decays, when $n \rightarrow \infty$, as

$$E^{k+p(n-p), p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{n-q < j \leq n-p} h_{ij}^\lambda \right) \simeq c_{p, q, k} n^{-(p^2-1)/2},$$

with

$$c_{p, q, k} := \frac{\sqrt{p} \prod_1^p (q - j)!}{(\sqrt{2\pi})^{p-1}} \binom{pq - 1 + k}{k}.$$

The next statement deals with the special case, where $q = p$. Namely, setting $N := n - p$, what is, asymptotically for large N , the probability $P^{Np+k, p}(\lambda_p \geq N)$?

Corollary 1.3. *Given an alphabet of size p and an integer $k > 0$, we give the behavior of the probability on the set of words of length $Np + k$ for large N . Notice $i_p(\pi) = \{\text{length of the word}\} = Np + k$ automatically, when $d_1(\pi) = p$. So, $i_{p-1}(\pi)$ is the first non-trivial quantity. We now have:*

$$\lim_{N \rightarrow \infty} N^{(p^2-1)/2} P^{Np+k, p}(\lambda_p \geq N)$$

$$= \lim_{N \rightarrow \infty} N^{(p^2-1)/2} P^{Np+k, p} \left(\begin{array}{c} d_1(\pi) = p \\ i_{p-1}(\pi) \leq N(p - 1) + k \end{array} \right)$$

$$= \frac{\sqrt{p} \prod_1^p (p - j)!}{(\sqrt{2\pi})^{p-1} k!} E \left(\left(\frac{1}{2} Z_{2p^2} \right)^k \right).$$

Thus the following decay holds for $N \nearrow \infty$,

$$P^{Np+k,p} \left(\begin{array}{l} d_1(\pi) = p \\ i_{p-1}(\pi) \leq N(p-1) + k \end{array} \right) \simeq c_{p,p,k} \left(\frac{1}{N} \right)^{(p^2-1)/2}.$$

The proofs of Theorem 1.2 and Corollary 1.3 will be given in section 4. In the next statement, we consider the expectation for the Poissonized probability $P_{x,p}$ on Young diagrams $\lambda \in \mathbb{Y}_\ell^{(p)}$, as defined in (1.0.4). Define

$$\mathcal{H}_p = \{p \times p \text{ Hermitian matrices}\},$$

and, for an interval $I \subset \mathbb{R}$,

$$\mathcal{H}_p(I) = \{M \in \mathcal{H}_p \text{ with spectrum in } I\}.$$

Theorem 1.4. Take $x > 0$, $s \in \mathbb{R}$ and set

$$(1.0.17) \quad n - p = x + s\sqrt{2x},$$

we have, upon expressing n in terms of x by means of the rescaling (1.0.17)³

$$\begin{aligned} (1.0.18) \quad & \lim_{x \rightarrow \infty} \frac{1}{(2x)^{p(q-p)/2}} E_{x,p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{ij}^\lambda \right) \\ &= \lim_{x \rightarrow \infty} \hat{c} e^{-px} x^{p(n-(p+q)/2)} \int_{\text{Gr}(p, \mathbb{C}^n)} e^{x \text{Tr}(I+Z^\dagger Z)^{-1}} \det(Z^\dagger Z)^{-(q-p)} d\rho(Z) \\ &= p! \frac{\int_{\mathcal{H}_p[s, \infty)} \det(M - sI)^{q-p} e^{-\text{Tr} M^2} dM}{\int_{\mathcal{H}_p} e^{-\text{Tr} M^2} dM} \\ &= c \exp \int_0^s h(y) dy, \end{aligned}$$

with $h(y)$ satisfying the **Painlevé IV** equation:

$$(1.0.19) \quad h''' + 6h'^2 - 4(y^2 + 2(q-2p))h' + 4yh - 8(q-p)p = 0.$$

The proof of Theorem 1.3 will be given in sections 3 and 6. In section 2 we show that the logarithmic derivative of a general multiple integral, involving the square of a Vandermonde, satisfies a third order differential equation, from which we derive, in

³ With $\hat{c} = \tilde{c} 2^{p(p-q)/2}$. Remember \tilde{c} from footnote 2.

section 3, that the logarithmic derivative of the matrix integral in (1.0.18) satisfies the Painlevé equation. Section 6 contains a discussion on how the Painlevé IV equation can be obtained from Painlevé V, by means of the rescaling (1.0.17), for large x . The following result shows that a certain general integral satisfies Painlevé V:

Theorem 1.5. *Given the weight*

$$\rho(z) = (z - a)^\alpha (b - z)^\beta e^{\gamma z}, \quad \text{with } \alpha, \beta > -1,$$

on the interval $[a, b]$, the integral

$$(1.0.20) \quad g(x) := \frac{\partial}{\partial x} \log \int_{[a, b]^n} \Delta^2(z) \prod_1^n e^{xz_k} \rho(z_k) dz_k$$

is a solution to the Painlevé V equation. To be precise,

$$(1.0.21) \quad f(y) = n(n + \alpha + \beta) - \frac{y}{b - a} \left(g\left(\frac{y}{b - a} - \gamma\right) - na \right)$$

satisfies a version of the Painlevé V equation,

$$(1.0.22) \quad f''^2 + \frac{4}{P^2} \left((Pf'^2 + Qf' + R)f' - (P'f'^2 + Q'f' + R')f \right. \\ \left. + \frac{1}{2}(P''f' + Q'')f^2 - \frac{1}{6}P'''f^3 - \frac{1}{4}\beta^2 n^2 \right) = 0;$$

with

$$P(y) = y$$

$$4Q(y) = -y^2 + 2y(2n + \alpha - \beta) - (\alpha + \beta)^2$$

$$2R(y) = -\beta n(\alpha + \beta + y).$$

Note that (1.0.22) is a well known form of Painlevé V, as discussed in the Appendix. Theorem 1.5 and also Theorem 3.1 below (which is exactly the last equality in (1.0.18)), can also be derived from the work of Forrester and Witte [12, 13].

2 A differential equation for a matrix integral

Proposition 2.1. *Consider a weight $\rho(z)$ on an interval $E \subseteq \mathbb{R}$, with rational logarithmic derivative of the form*

$$(2.0.1) \quad -\frac{\rho'(z)}{\rho(z)} = \frac{b_0 + b_1 z + b_2 z^2}{a_0 + a_1 z + a_2 z^2} =: \frac{B(z)}{A(z)},$$

and boundary condition

$$(2.0.2) \quad A(z)\rho(z)z^k|_{\partial E} = 0, \quad \text{for all } k = 0, 1, 2, \dots$$

The expression⁴

$$(2.0.3) \quad g(x) := \frac{\partial}{\partial x} \log \int_{E^n} \Delta^2(z) \prod_1^n e^{xz_k} \rho(z_k) dz_k$$

satisfies a third order differential equation, namely:

$$(2.0.4) \quad g''' + 6g'^2 + \frac{4a_2(g'' + 2gg')}{a_2x - b_2} + \frac{2a_2^2g^2 + P_2g'}{(a_2x - b_2)^2} + \frac{P_1g - nQ_1}{(a_2x - b_2)^3} = 0,$$

which can be transformed into the Painlevé V equation. In (2.0.4), the P_i 's and Q_1 are polynomials in x ,

$$\begin{aligned} P_2(x) &:= \left\{ \begin{aligned} &(4a_0a_2 - a_1^2)x^2 + 2(2a_1a_2n - 2a_0b_2 + a_1b_1 - 2a_2b_0)x \\ &- 4a_2^2n^2 - 4(2a_1b_2 - a_2b_1)n + 4b_0b_2 - b_1^2 + 2a_2^2 \end{aligned} \right\} \\ P_1(x) &:= \left\{ \begin{aligned} &(2a_1a_2^2n + 2a_0a_2b_2 - a_1^2b_2 + a_1a_2b_1 - 2a_2^2b_0)x \\ &- 4a_2^3n^2 + (-6a_1a_2b_2 + 4a_2^2b_1)n \\ &- 2a_0b_2^2 + a_1b_1b_2 + 2a_2b_0b_2 - a_2b_1^2 \end{aligned} \right\} \\ Q_1(x) &:= \left\{ \begin{aligned} &(2a_0a_2^2n - a_1^2a_2n + a_0a_1b_2 - 2a_0a_2b_1 + a_1a_2b_0)x \\ &+ 2a_1a_2^2n^2 + (2a_1^2b_2 - a_1a_2b_1 - 2a_2^2b_0)n \\ &+ a_0b_1b_2 - 2a_1b_0b_2 + a_2b_0b_1 \end{aligned} \right\}. \end{aligned}$$

The proof of Proposition 2.1 hinges on the following lemma, which we state in its full generality, although only the case $I = E$ will be used; see [1, 23].

Lemma 2.2. *Given a disjoint union of intervals $I = \bigcup_1^r [c_{2i-1}, c_{2i}] \subset E$, the integral*

$$\tau_n(t, c) = \int_{I^n} \Delta^2(z) \prod_1^n e^{\sum_{i=1}^\infty t_i z_k^i} \rho(z_k) dz_k$$

with $\rho(z)$ and E as in (2.0.1) and (2.0.2) satisfies

⁴ When E is a finite interval, the integral always converges, and for a infinite interval, one may have to require $x > \alpha$ or $x < \alpha$, for some $\alpha \in \mathbb{R}$.

(i) Virasoro constraints for all $m \geq -1$:

$$(2.0.5) \quad \left(-\sum_1^{2r} c_i^{m+1} A(c_i) \frac{\partial}{\partial c_i} + \mathbb{V}_m \right) \tau_n(t, c) = 0,$$

with

$$\mathbb{V}_m := \sum_{k=0}^2 \left\{ \begin{array}{l} a_k (J_{k+m}^{(2)} + 2nJ_{k+m}^{(1)} + n^2 J_{k+m}^{(0)}) \\ - b_k (J_{k+m+1}^{(1)} + n\delta J_{k+m+1}^{(0)}) \end{array} \right\},$$

where

$$J_k^{(2)} = \sum_{i+j=k} \frac{\partial^2}{\partial t_i \partial t_j} + \sum_{-i+j=k} it_i \frac{\partial}{\partial t_j} + \frac{1}{4} \sum_{-i-j=k} it_i jt_j$$

$$J_k^{(1)} = \frac{\partial}{\partial t_k} + \frac{1}{2}(-k)t_k, \quad J_k^{(0)} = \delta_{k0}.$$

(ii) The KP-hierarchy⁵ ($k = 0, 1, 2, \dots$)

$$\left(\mathbf{s}_{k+4} \left(\frac{\partial}{\partial t_1}, \frac{1}{2} \frac{\partial}{\partial t_2}, \frac{1}{3} \frac{\partial}{\partial t_3}, \dots \right) - \frac{1}{2} \frac{\partial^2}{\partial t_1 \partial t_{k+3}} \right) \tau_n \circ \tau_n = 0,$$

of which the first equation reads:

$$(2.0.6) \quad \left(\left(\frac{\partial}{\partial t_1} \right)^4 + 3 \left(\frac{\partial}{\partial t_2} \right)^2 - 4 \frac{\partial^2}{\partial t_1 \partial t_3} \right) \log \tau_n + 6 \left(\frac{\partial^2}{\partial t_1^2} \log \tau_n \right)^2 = 0.$$

Proof. The most transparent way to prove this lemma is via vector vertex operators, for which the β -integrals

$$(2.0.7) \quad \tau_n(t, c; \beta) := \int_{E^n} |\Delta_n(x)|^{2\beta} \prod_{k=1}^n (e^{\sum_1^\infty t_i x_k^i} \rho(x_k) dx_k), \quad \text{for } n > 0$$

are fixed points (see [1]). Another method, more computational, but much less conceptual, is to use a self-similarity argument, as in [1]. Namely, setting

⁵ Given a polynomial $p(t_1, t_2, \dots)$, define the customary Hirota symbol $p(\partial_t) f \circ g := p\left(\frac{\partial}{\partial y_1}, \frac{\partial}{\partial y_2}, \dots\right) f(t+y)g(t-y)|_{y=0}$. The \mathbf{s}_ℓ 's are the elementary Schur polynomials $e^{\sum_1^\infty t_i z^i} := \sum_{i \geq 0} \mathbf{s}_i(t) z^i$ and for later use, set $\mathbf{s}_\ell(\tilde{\partial}) := \mathbf{s}_\ell\left(\frac{\partial}{\partial t_1}, \frac{1}{2} \frac{\partial}{\partial t_2}, \dots\right)$.

$$d\tau_n(z) := |\Delta_n(z)|^{2\beta} \prod_{k=1}^n (e^{\sum_{i=1}^{\infty} t_i z_k^i} \rho(z_k) dz_k),$$

we have the following variational formula:

$$(2.0.8) \quad \frac{d}{d\varepsilon} d\tau_n(z_i \mapsto z_i + \varepsilon A(z_i) z_i^{k+1})|_{\varepsilon=0} = \sum_{\ell=0}^2 (a_{\ell}^{\beta} \mathbb{J}_{k+\ell, n}^{(2)} - b_{\ell}^{\beta} \mathbb{J}_{k+\ell+1, n}^{(1)}) d\tau_n,$$

with

$$\begin{aligned} \beta \mathbb{J}_{k, n}^{(2)}(t, n) &= \beta J_k^{(2)} + (2n\beta + (k+1)(1-\beta)) \beta J_k^{(1)} + n((n-1)\beta + 1) \delta_{k0}, \\ \beta \mathbb{J}_{k, n}^{(1)}(t, n) &= \beta J_k^{(1)} + n \delta_{k0}, \end{aligned}$$

where

$$(2.0.9) \quad \begin{aligned} \beta J_k^{(1)} &= \frac{\partial}{\partial t_k} + \frac{1}{2\beta} (-k) t_{-k} \\ \beta J_k^{(2)} &= \sum_{i+j=k} \frac{\partial^2}{\partial t_i \partial t_j} + \frac{1}{\beta} \sum_{-i+j=k} i t_i \frac{\partial}{\partial t_j} + \frac{1}{4\beta^2} \sum_{-i-j=k} i t_i j t_j. \end{aligned}$$

The change of integration variable $z_i \mapsto z_i + \varepsilon A(z_i) z_i^{k+1}$ in the integral (2.0.7) leaves the integral invariant, but it induces a change of limits of integration, given by the inverse of the map above; namely the c_i 's in $I = \bigcup_1^r [c_{2i-1}, c_{2i}]$, get mapped as follows

$$c_i \mapsto c_i - \varepsilon A(c_i) c_i^{k+1} + O(\varepsilon^2).$$

Therefore, setting

$$I^{\varepsilon} = \bigcup_1^r [c_{2i-1} - \varepsilon A(c_{2i-1}) c_{2i-1}^{k+1} + O(\varepsilon^2), c_{2i} - \varepsilon A(c_{2i}) c_{2i}^{k+1} + O(\varepsilon^2)],$$

we find, using (2.0.8) and the fundamental theorem of calculus,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \varepsilon} \int_{(I^{\varepsilon})^{2n}} |\Delta_{2n}(z + \varepsilon A(z) z^{k+1})|^{2\beta} \prod_{i=1}^n e^{-V(z_i + \varepsilon A(z_i) z_i^{k+1}, t)} d(z_i + \varepsilon A(z_i) z_i^{k+1}) \\ &= \left(-\sum_{i=1}^{2r} c_i^{k+1} A(c_i) \frac{\partial}{\partial c_i} + \sum_{\ell=0}^2 (a_{\ell}^{\beta} \mathbb{J}_{k+\ell, 2n}^{(2)} - b_{\ell}^{\beta} \mathbb{J}_{k+\ell+1, 2n}^{(1)}) \right) \tau_n(t, c, \beta). \end{aligned}$$

For Lemma 2.2 one sets $\beta = 1$; also, when $I = E$, the condition (2.0.2) implies that the boundary terms in the formula above are absent. For statement (ii), concerning the KP equation, we refer the reader to [23]. This ends the proof of Lemma 2.2. \square

Proof of Proposition 2.1. Setting $F(t) := F_n(t) = \log \tau_n(t)$, a few of the Virasoro constraints of Lemma 2.2, evaluated along the locus

$$\mathcal{L} := \{t = (x, 0, 0, \dots)\},$$

read as follows:

$$\begin{aligned} \left. \frac{\mathbb{V}_{-1}\tau_n}{\tau_n} \right|_{\mathcal{L}} &= a_0 \left(\sum_{i \geq 2} it_i \frac{\partial F}{\partial t_{i-1}} + nt_1 \right) + a_1 \left(\sum_{i \geq 1} it_i \frac{\partial F}{\partial t_i} + n^2 \right) \\ &\quad + a_2 \left(\sum_{i \geq 1} it_i \frac{\partial F}{\partial t_{i+1}} + 2n \frac{\partial F}{\partial t_1} \right) - \left(b_0 n + b_1 \frac{\partial F}{\partial t_1} + b_2 \frac{\partial F}{\partial t_2} \right) \Big|_{\mathcal{L}} \\ &= n(a_0 x + a_1 n - b_0) + (a_1 x + 2na_2 - b_1) \frac{\partial F}{\partial t_1} + (a_2 x - b_2) \frac{\partial F}{\partial t_2} \Big|_{\mathcal{L}} = 0. \\ \left. \frac{\mathbb{V}_0\tau_n}{\tau_n} \right|_{\mathcal{L}} &= a_0 \left(\sum_{i \geq 1} it_i \frac{\partial F}{\partial t_i} + n^2 \right) + a_1 \left(\sum_{i \geq 1} it_i \frac{\partial F}{\partial t_{i+1}} + 2n \frac{\partial F}{\partial t_1} \right) \\ &\quad + a_2 \left(\sum_{i \geq 1} it_i \frac{\partial F}{\partial t_{i+2}} + \frac{\partial^2 F}{\partial t_1^2} + \left(\frac{\partial F}{\partial t_1} \right)^2 + 2n \frac{\partial F}{\partial t_2} \right) \\ &\quad - \left(b_0 \frac{\partial F}{\partial t_1} + b_1 \frac{\partial F}{\partial t_2} + b_2 \frac{\partial F}{\partial t_3} \right) \Big|_{\mathcal{L}} \\ &= a_0 n^2 + (a_0 x + 2na_1 - b_0) \frac{\partial F}{\partial t_1} + a_2 \left(\frac{\partial F}{\partial t_1} \right)^2 + a_2 \frac{\partial^2 F}{\partial t_1^2} \\ &\quad + (a_1 x + 2na_2 - b_1) \frac{\partial F}{\partial t_2} + (a_2 x - b_2) \frac{\partial F}{\partial t_3} \Big|_{\mathcal{L}} = 0. \\ \left. \frac{\partial}{\partial t_1} \frac{\mathbb{V}_{-1}\tau_n}{\tau_n} \right|_{\mathcal{L}} &= a_0 n + \sum_1^2 a_i \frac{\partial F}{\partial t_i} + (a_1 x + 2na_2 - b_1) \frac{\partial^2 F}{\partial t_1^2} \\ &\quad + (a_2 x - b_2) \frac{\partial^2 F}{\partial t_1 \partial t_2} \Big|_{\mathcal{L}} = 0. \end{aligned}$$

$$\begin{aligned}
 \left. \frac{\partial}{\partial t_1} \frac{\mathbf{V}_0 \tau_n}{\tau_n} \right|_{\mathcal{L}} &= \sum_{i=0}^2 a_i \frac{\partial F}{\partial t_{i+1}} + a_2 \left(\frac{\partial^3 F}{\partial t_1^3} + 2 \frac{\partial F}{\partial t_1} \frac{\partial^2 F}{\partial t_1^2} \right) \\
 &\quad + (a_0 x + 2na_1 - b_0) \frac{\partial^2 F}{\partial t_1^2} + (a_1 x + 2n - b_1) \frac{\partial^2 F}{\partial t_1 \partial t_2} \\
 &\quad + (a_2 x - b_2) \frac{\partial^2 F}{\partial t_1 \partial t_3} \Big|_{\mathcal{L}} = 0. \\
 \\
 \left. \frac{\partial}{\partial t_2} \frac{\mathbf{V}_{-1} \tau_n}{\tau_n} \right|_{\mathcal{L}} &= 2 \sum_{i=0}^2 a_i \frac{\partial F}{\partial t_{i+1}} + (a_1 x + 2na_2 - b_1) \frac{\partial^2 F}{\partial t_1 \partial t_2} \\
 &\quad + (a_2 x - b_2) \frac{\partial^2 F}{\partial t_2^2} \Big|_{\mathcal{L}} = 0.
 \end{aligned}$$

These five equations form a linear system in the five unknowns

$$\left. \frac{\partial F}{\partial t_2} \right|_{\mathcal{L}}, \quad \left. \frac{\partial^2 F}{\partial t_1 \partial t_2} \right|_{\mathcal{L}}, \quad \left. \frac{\partial^2 F}{\partial t_2^2} \right|_{\mathcal{L}}, \quad \left. \frac{\partial F}{\partial t_3} \right|_{\mathcal{L}}, \quad \left. \frac{\partial^2 F}{\partial t_1 \partial t_3} \right|_{\mathcal{L}},$$

which upon solving in terms of $\left(\frac{\partial}{\partial t_1}\right)^k F|_{\mathcal{L}}$ and substituting into the KP-equation (remembering equation (2.0.6))

$$\left(\frac{\partial^4}{\partial t_1^4} + 3 \frac{\partial^2}{\partial t_2^2} - 4 \frac{\partial^2}{\partial t_1 \partial t_3} \right) F + 6 \left(\frac{\partial^2 F}{\partial t_1^2} \right)^2 = 0$$

yields a differential equation in F . Pure F never appears in this equation, because the Virasoro constraints only contain partials of F . Therefore, it is a differential equation in $g(x) = \frac{\partial}{\partial t_1} F(t_1, 0, 0, \dots)|_{t_1=x}$, which one computes has the form (2.0.4); this ends the proof of Proposition 2.1. \square

The proof of Theorem 1.5 will be given at the end of section 3.

3 Hermitian matrix integrals and Painlevé equations

Define

$$\mathcal{H}_n = \{n \times n \text{ Hermitian matrices}\},$$

and, for an interval $I \subset \mathbb{R}$,

$$\mathcal{H}_n(I) = \{M \in \mathcal{H}_n \text{ with spectrum in } I\}.$$

We state the following theorem, which, as we found out, was already in the work of Forrester and Witte [12, 13]; our method using Proposition 2.1 is very different.

Theorem 3.1. *For⁶ $a, b > -1$, the logarithmic derivatives of the integrals*

$$(3.0.1) \quad h(s) := \frac{d}{ds} \log \int_{\mathcal{H}_n(-\infty, s]} \det(M - sI)^a e^{-\text{Tr } M^2} dM$$

$$k(s) := s \frac{d}{ds} \log \int_{\mathcal{H}_n[0, s]} \det(sI - M)^b \det M^a e^{-\text{Tr } M} dM$$

satisfy the Painlevé IV and V equations, respectively:

$$(3.0.2) \quad h''' + 6h'^2 - 4(s^2 + 2(a - n))h' + 4sh - 8an = 0,$$

and

$$(3.0.3) \quad k''' + \frac{k''}{s} + \frac{6}{s}k'^2 - \frac{4}{s^2}kk' - (s^2 - 2s(2n + a - b) + (a + b)^2) \frac{k'}{s^2}$$

$$+ (s - 2n - a + b) \frac{k}{s^2} - \frac{bn}{s^2}(s + a + b) = 0.$$

Proof of Theorem 3.1. Set

$$(3.0.4) \quad \tau(x) = \int_{I^n} \Delta^2(z) \prod_1^n e^{xz_k} \rho(z_k) dz_k,$$

with ρ and I as in (2.0.1) and (2.0.2).

(i) Then, expressed in spectral coordinates and making the substitution $y_i = z_i - s$ for $1 \leq i \leq n$, the first matrix integral in (3.0.1) reads, for $I = (-\infty, s], [s, \infty)$ and $(-\infty, \infty)$,

$$(3.0.5) \quad \int_{(I_1)^n} \Delta^2(z) \prod_1^n (z_i - s)^a e^{-z_i^2} dz_i$$

$$= e^{-ns^2} \int_{(I_1')^n} \Delta^2(y) \prod_1^n y_i^a e^{-y_i^2 - 2sy_i} dy_i = e^{-ns^2} \tau(-2s)$$

⁶ For the first integral of (3.0.1), one can choose the intervals $I_1 = (-\infty, s], [s, \infty)$ or $(-\infty, \infty)$. For the second integral of (3.0.1), one may choose the intervals $I_2 = [0, s], [s, \infty)$ or $[0, \infty)$.

with $I'_1 = (-\infty, 0], [0, \infty)$ or $(-\infty, \infty)$. Here $\tau(x)$ as in (3.0.4) contains $\rho(z) = z^a e^{-z^2}$, for which

$$-\frac{\rho'}{\rho} = \frac{-a + 2z^2}{z}.$$

Thus, setting

$$(3.0.6) \quad a_0 = a_2 = 0, \quad a_1 = 1$$

$$b_0 = -a, \quad b_1 = 0, \quad b_2 = 2,$$

in the equation (2.0.4), one deduces that

$$g(x) := \frac{\partial}{\partial x} \log \tau(x), \quad \text{for } I = \begin{cases} (-\infty, 0] \\ (-\infty, \infty) \\ (0, \infty) \end{cases},$$

satisfies

$$g''' + 6g'^2 - g' \left(\frac{x^2}{4} + 4n + 2a \right) + \frac{xg}{4} + \frac{n}{2}(n+a) = 0.$$

But we need a differential equation for $e^{-ns^2} \tau(-2s)$, instead of $\tau(x)$. Therefore consider

$$h(s) = \frac{\partial}{\partial s} \log(e^{-ns^2} \tau(-2s)) = -2ns - 2g(x)|_{x=-2s},$$

which relates to $g(x)$ as follows,

$$h'(s) = -2n + 4g'(x)|_{x=-2s}$$

$$h''(s) = -8g''(x)|_{x=-2s}$$

$$h'''(s) = 16g'''(x)|_{x=-2s}.$$

Expressing $g(x)$, $g'(x)$, $g''(x)$, $g'''(x)|_{x=-2s}$ in terms of h , h' , h'' and h''' , and setting $x = -2s$ yield the differential equation (3.0.2), which according to the table in Appendix 1 is a version of Painlevé IV; this establishes the first part of Theorem 3.1.

(ii) Then, making the substitution $y_i = z_i/s$ for $1 \leq i \leq n$,

$$\begin{aligned}
 (3.0.7) \quad & \int_{(I_2)^n} \Delta^2(z) \prod_1^n (s - z_i)^b z_i^a e^{-z_i} dz_i \\
 &= s^{n(n+a+b)} \int_{(I'_2)^n} \Delta^2(y) \prod_1^n (1 - y_i)^b y_i^a e^{-sy_i} dy_i \\
 &= s^{n(n+a+b)} \tau(-s)
 \end{aligned}$$

with $I_2 = [0, s]$, $I'_2 = [0, 1]$, for $I_2 = [s, \infty]$, $I'_2 = [1, \infty]$ and finally for $I_2 = [0, \infty]$, $I'_2 = [0, \infty]$; $\tau(x)$ now corresponds to $\rho(z) = z^a(1 - z)^b$, and so

$$-\frac{\rho'}{\rho} = -\frac{a}{z} + \frac{b}{1-z} = \frac{a - (a+b)z}{-z + z^2}.$$

Thus

$$\begin{aligned}
 a_0 &= 0, & a_1 &= -1, & a_2 &= 1 \\
 b_0 &= a, & b_1 &= -(a+b), & b_2 &= 0.
 \end{aligned}$$

Setting these special values in the equation (2.0.4), one checks that

$$g(x) := \frac{\partial}{\partial x} \log \int_{I_2^m} \Delta^2(z) \prod_1^n e^{xz_k} z_k^a (1 - z_k)^b dz_k,$$

satisfies

$$\begin{aligned}
 (3.0.8) \quad & g''' + 6g'^2 + \frac{4}{x}(g'' + 2gg') + 2\left(\frac{g}{x}\right)^2 \\
 & - \frac{g'}{x^2}(x^2 + 2(2n + a - b)x + (2n + a + b)^2 - 2) \\
 & - \frac{g}{x^3}((2n + a - b)x + (2n + a + b)^2) \\
 & + \frac{n}{x^3}(n + a)(x + 2n + a + b) = 0.
 \end{aligned}$$

Since $g(x) = \frac{\partial}{\partial x} \log \tau(x)$, we have

$$k(s) = s \frac{\partial}{\partial s} \log s^{n(n+a+b)} \tau(-s) = n(n + a + b) - sg(-s)$$

and so

$$g(-s) = \frac{1}{s}(-k(s) + n(n + a + b))$$

$$g'(-s) = \frac{1}{s^2}(sk'(s) - k(s) + n(n + a + b))$$

$$g''(-s) = \frac{2}{s^3}(-s^2k''(s)/2 + sk'(s) - k(s) + n(n + a + b))$$

$$g'''(-s) = \frac{3}{s^4}(s^3k'''(s)/3 - s^2k''(s) + 2sk'(s) - 2k(s) + 2n(n + a + b)).$$

Substituting these expressions into (3.0.8) yields the differential equation (3.0.3), which again referring to the table in appendix 1 is Painlevé V, ending the proof of Theorem 3.1. \square

Proof of Theorem 1.5. Because of the form (2.0.1) of $\rho(z) dz$, we have for an open set of real constants a_i and b_i ,

$$\rho(z) = (z - a)^\alpha (b - z)^\beta e^{\gamma z}, \quad \text{with } \alpha, \beta > -1 \text{ and } a, b \in \mathbb{R},$$

which, upon making a linear change of variables $z \mapsto y$ in the integral (2.0.3), leads to

$$g(x) = na + \frac{\partial}{\partial x} \log \int_{[0,1]^n} \Delta^2(y) \prod_1^n e^{(x+\gamma)(b-a)y_k} y_k^\alpha (1 - y_k)^\beta dy_k$$

and so

$$\frac{1}{b-a} \left(g \left(\frac{x'}{b-a} - \gamma \right) - na \right) = \frac{\partial}{\partial x'} \log \int_{[0,1]^n} \Delta^2(y) \prod_1^n e^{x' y_k} y_k^\alpha (1 - y_k)^\beta dy_k,$$

which is shown to be a solution of Painlevé V in part (ii) of the proof of Theorem 3.1 in this section.

To compute the constant in (1.0.22) (that is δ in equation (7.0.2)), multiply the equation (1.0.22) with $P^2/4$, set $y = 0$ and use the explicit expressions for the polynomials P, Q, R and the function $f(y)$, as in (1.0.21), yielding the identity

$$(3.0.9) \quad \delta = -[Qf' + R]f' + [f'^2 + Q'f' + R']f - \frac{1}{2}Q''f^2|_{y=0}.$$

All the expressions can readily be computed, except for $f'(0)$, which requires some argument. One computes

$$\begin{aligned}
 f'(0) &= \frac{1}{b-a}(na - g(-\gamma)) \\
 &= \frac{n}{b-a} \left(a - \frac{\int_{[a,b]^n} \Delta^2(z) \left(\frac{1}{n} \sum_1^n z_k\right) (z_k - a)^\alpha (b - z_k)^\beta dz_k}{\int_{[a,b]^n} \Delta^2(z) \prod_1^n (z_k - a)^\alpha (b - z_k)^\beta dz_k} \right) \\
 &= \frac{n}{b-a} (a - \langle z_1 \rangle_{(a,b)}) \\
 &= -\frac{n(n+\alpha)}{2n+\alpha+\beta}.
 \end{aligned}$$

To establish the last equality above, we need the Aomoto extension [5] (see [2], Appendix D) of Selberg's integral:

$$\begin{aligned}
 (3.0.10) \quad \langle x_1 \dots x_m \rangle_{(0,1)} &:= \frac{\int_{[0,1]^n} x_1 \dots x_m |\Delta(x)|^{2\gamma} \prod_{j=1}^n x_j^\alpha (1-x_j)^\beta dx_j}{\int_{[0,1]^n} |\Delta(x)|^{2\gamma} \prod_{j=1}^n x_j^\alpha (1-x_j)^\beta dx_j} \\
 &= \prod_{j=1}^m \frac{\alpha + 1 + (n-j)\gamma}{\alpha + \beta + 2 + (2n-j-1)\gamma}.
 \end{aligned}$$

In particular, setting $\gamma = 1$, this formula implies

$$\begin{aligned}
 (3.0.11) \quad \langle z_1 \rangle_{(a,b)} &= \langle a + (b-a)x_1 \rangle_{(a,b)} \\
 &= a + (b-a) \langle x_1 \rangle_{(0,1)} = a + (b-a) \frac{n+\alpha}{2n+\beta+\alpha}
 \end{aligned}$$

Adding up all the pieces in (3.0.9), one finds the value of the constant δ in equation (7.0.2), ending the proof of Theorem 1.5. \square

4 A limit theorem for the probability on words and Chi-square

The purpose of this section is to prove Theorem 1.2 and Corollary 1.3. Given m independent normal $N(0, 1)$ -random variables X_i , the sum

$$Z_m = \sum_1^m X_i^2$$

is χ_m^2 -distributed, with Fourier transform

$$E(e^{tZ_m}) = (1 - 2t)^{-m/2}.$$

Developing both sides in t yields the moments

$$\frac{1}{k!} E \left(\left(\frac{1}{2} Z_m \right)^k \right) = \binom{\frac{m}{2} - 1 + k}{k}.$$

Proof of Theorem 1.2. From Stirling formula $N! = \sqrt{2\pi N} e^{N(\log N - 1)} (1 + \mathbf{O}(\frac{1}{N}))$ and $(N + \alpha)! \simeq N! N^\alpha$, we have for large n ,

$$(4.0.1) \quad (pn + k - p^2)! \simeq (pn)!(pn)^{k-p^2} \simeq \frac{(n!p^n)^p}{(2\pi n)^{(p-1)/2}} \sqrt{p} (pn)^{k-p^2}$$

$$\prod_1^p (n - j)! \simeq \prod_1^p (n! n^{-j}) \simeq (n!)^p n^{-p(p+1)/2}$$

meaning here that the ratios between two consecutive expressions tend to 1, when $n \rightarrow \infty$. On the one hand, we have, changing the summation index $\ell \rightarrow \ell - p(n - p)$,

$$\begin{aligned} (4.0.2) \quad & \lim_{n \rightarrow \infty} \prod_{j=1}^p \frac{(n-j)!}{(q-j)!} \sum_{\ell \geq p(n-p)} \frac{p^\ell (nu)^{\ell-p(n-p)}}{\ell!} E^{\ell,p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{n-q < j \leq n-p} h_{ij}^\lambda \right) \\ &= \lim_{n \rightarrow \infty} \prod_{j=1}^p \frac{(n-j)!}{(q-j)!} \sum_{k=0}^{\infty} \frac{p^{p(n-p)} (pnu)^k}{(k + p(n-p))!} E^{k+p(n-p),p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{n-q < j \leq n-p} h_{ij}^\lambda \right) \\ &= \sum_{k=0}^{\infty} u^k \lim_{n \rightarrow \infty} \frac{(pn)^k p^{p(n-p)}}{\prod_1^p (q-j)!} \frac{\prod_1^p (n-j)!}{(pn + k - p^2)!} E^{k+p(n-p),p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{n-q < j \leq n-p} h_{ij}^\lambda \right) \\ &= \sum_{k=0}^{\infty} u^k \lim_{n \rightarrow \infty} \frac{(2\pi)^{(p-1)/2} n^{(p^2-1)/2}}{\prod_1^p (q-j)! \sqrt{p}} E^{k+p(n-p),p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{n-q < j \leq n-p} h_{ij}^\lambda \right), \end{aligned}$$

using (4.0.1) in the last equality.

On the other hand, using (1.0.11), (1.0.5), (1.0.13) and

$$(4.0.3) \quad \lim_{n \rightarrow \infty} \frac{(n)_\lambda}{n^{|\lambda|}} = \lim_{n \rightarrow \infty} \left[\frac{n(n+1) \dots (n+\lambda_1-1)}{n^{\lambda_1}} \right] \left[\frac{(n-1)n \dots (n+\lambda_2-2)}{n^{\lambda_2}} \right] \dots,$$

$$= 1$$

we have

$$\begin{aligned}
(4.0.4) \quad & \lim_{n \rightarrow \infty} \prod_{j=1}^p \frac{(n-j)!}{(q-j)!} \sum_{\ell \geq p(n-p)} \frac{p^\ell (nu)^{\ell-p(n-p)}}{\ell!} E^{\ell,p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{n-q < j \leq n-p} h_{ij}^\lambda \right) \\
&= \lim_{n \rightarrow \infty} {}_2F_1^{(1)}(p, q; n; x) |_{\sum_\ell x_\ell^i = nu \delta_{i1}} \\
&= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} u^k \sum_{\kappa \in \mathbb{Y}_k} \frac{n^\kappa(p)_\kappa(q)_\kappa}{(h^\kappa)^2(n)_\kappa} \\
&= \sum_{k=0}^{\infty} u^k \sum_{\kappa \in \mathbb{Y}_k} \frac{(p)_\kappa(q)_\kappa}{(h^\kappa)^2} \\
&= \sum_{k=0}^{\infty} u^k \sum_{\kappa \in \mathbb{Y}_k} \mathbf{s}_\kappa(1^p) \mathbf{s}_\kappa(1^q), \quad \text{using (1.0.7)} \\
&= \sum_{\kappa \in \mathbb{Y}} \mathbf{s}_\kappa(u^p) \mathbf{s}_\kappa(1^q) \\
&= (1-u)^{-pq}, \quad \text{using the Cauchy identity} \\
&= E(e^{(1/2)uZ_{2pq}}).
\end{aligned}$$

Then comparing the coefficients of u^k in the identical asymptotic expressions (4.0.2) and (4.0.4) yields Theorem 1.2. \square

Proof of Corollary 1.3. Setting $p = q$ and $N = n - p$ in Theorem 1.2, the expectation in (1.0.16) becomes $P^{Np+k,p}(\lambda \supseteq \mu)$ for fixed μ . By RSK and (1.0.9), the condition $\lambda \supseteq \mu$ translates into $d_1(\pi) = \lambda_1^\top = p$ and $i_{p-1}(\pi) = \sum_1^p \lambda_i - \lambda_p = Np + k - \lambda_p$, with $\lambda_p \geq N$. This means $i_{p-1}(\pi) \leq k + (p-1)N$. \square

5 A limit theorem for the Poissonized probability on words

In the next theorem, we consider the expectation for the Poissonized probability $P_{x,p}$ on Young diagrams $\lambda \in \mathbb{Y}_\ell^{(p)}$, as defined in (1.0.4). Before proving Theorem 1.4, one needs the following proposition:

Proposition 5.1 ([15, 22]). *For every continuous function g on \mathbb{R}^p , we have*

$$\begin{aligned}
& \lim_{\ell \rightarrow \infty} E^{\ell,p} \left(g \left(\frac{\lambda_1 - \ell/p}{\sqrt{2\ell/p}}, \dots, \frac{\lambda_p - \ell/p}{\sqrt{2\ell/p}} \right) \right) \quad |\lambda| = \ell \\
&= p! \sqrt{\pi p} \int_{\substack{x \in \mathbb{R}^p \\ x_1 > \dots > x_p \\ \sum_{i=1}^p x_i = 0}} g(x_1, \dots, x_p) \varphi_p(x_1, \dots, x_p) dx_1 \dots dx_{p-1}
\end{aligned}$$

and

$$\begin{aligned} & \lim_{x \rightarrow \infty} E_{x,p} \left(g \left(\frac{\lambda_1 - x}{\sqrt{2x}}, \dots, \frac{\lambda_p - x}{\sqrt{2x}} \right) \right) \\ &= p! \int_{\substack{x \in \mathbb{R}^p \\ x_1 > \dots > x_p}} g(x_1, \dots, x_p) \varphi_p(x_1, \dots, x_p) dx_1 \dots dx_p, \end{aligned}$$

where $\varphi_p(x_1, \dots, x_p)$ is the probability density

$$\varphi_p(x_1, \dots, x_p) = \frac{1}{Z_p} \Delta_p(x)^2 \prod_{j=1}^p e^{-x_j^2},$$

with

$$Z_p = (2\pi)^{p/2} \frac{2^{-p^2/2}}{\prod_1^p j!}.$$

Proof of Theorem 1.4. Given partition λ with the probability measure (1.0.4), consider the random variable

$$(5.0.1) \quad \varepsilon_i(\lambda) := \frac{\lambda_i - x}{\sqrt{2x}}.$$

From (1.0.17), we have

$$(5.0.2) \quad s = \frac{(n-p) - x}{\sqrt{2x}}.$$

Consider the hook length $h_{(ij)}^\lambda$ for box $(i, j) \in \lambda$, such that $n - q < j \leq n - p$. Define r_{ij} such that

$$\begin{aligned} h_{(ij)}^\lambda &:= \lambda_i - (n - p) + r_{ij} \\ &= \sqrt{2x}(\varepsilon_i - s) + r_{ij}, \end{aligned}$$

upon using formulas (5.0.1) and (5.0.2) in the last equality. The position (i, j) in the partition λ such that $n - q < j \leq n - p$ implies that $r_{ij} \leq (q - p) + p = q$, from visual inspection of Figure 1. We also have

$$\begin{aligned}
\{\lambda \in \mathbb{Y}^{(p)}, \lambda \supseteq \mu\} &= \{\lambda \in \mathbb{Y}^{(p)}, \lambda_i \geq n-p, \text{ all } 1 \leq i \leq p\} \\
&= \left\{ \lambda \in \mathbb{Y}^{(p)}, \frac{\lambda_i - x}{\sqrt{2x}} \geq \frac{(n-p) - x}{\sqrt{2x}}, \text{ all } 1 \leq i \leq p \right\} \\
&= \bigcap_{i=1}^p \{\lambda \in \mathbb{Y}^{(p)}, \varepsilon_i(\lambda) \geq s\}.
\end{aligned}$$

On the set $\varepsilon_i(\lambda) \geq s$ and for $\lambda \in \mathbb{Y}^{(p)}$, we have, since $0 \leq r_{ij} \leq q$,

$$\begin{aligned}
&(2x)^{-p(q-p)/2} \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda - \prod_{1 \leq i \leq p} (\varepsilon_i - s)^{q-p} \\
&= (2x)^{-p(q-p)/2} \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} (\sqrt{2x}(\varepsilon_i - s) + r_{ij}) - \prod_{1 \leq i \leq p} (\varepsilon_i - s)^{q-p} \\
&= \prod_{\substack{1 \leq i \leq p \\ n-q < j \leq n-p}} \left((\varepsilon_i - s) + \frac{r_{ij}}{\sqrt{2x}} \right) - \prod_{1 \leq i \leq p} (\varepsilon_i - s)^{q-p} \\
&\leq \prod_{1 \leq i \leq p} \left((\varepsilon_i - s) + \frac{q}{\sqrt{2x}} \right)^{q-p} - \prod_{1 \leq i \leq p} (\varepsilon_i - s)^{q-p} \\
&= \sum_{\substack{0 \leq \ell_i \leq q-p \\ 1 \leq i \leq p \\ \sum \ell_i \geq 1}} \left(\frac{q}{\sqrt{2x}} \right)^{\sum \ell_i} \prod_{i=1}^p \binom{q-p}{\ell_i} \prod_{i=1}^p (\varepsilon_i - s)^{q-p-\ell_i}.
\end{aligned}$$

The positive expression is now estimated as follows:

$$\begin{aligned}
&(2x)^{-p(q-p)/2} E^{\ell,p} \left(I_{\{\lambda \supseteq \mu\}}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) \\
&\quad - E^{\ell,p} \left(\prod_{1 \leq i \leq p} (\varepsilon_i - s)^{q-p} I_{\varepsilon_i \geq s} \right) \\
&\leq \sum_{\substack{0 \leq k_i \leq q-p \\ 1 \leq i \leq p \\ \sum k_i \geq 1}} \left(\frac{q}{\sqrt{2x}} \right)^{\sum k_i} \prod_{i=1}^p \binom{q-p}{k_i} E^{\ell,p} \prod_{i=1}^p (\varepsilon_i - s)^{q-p-k_i} I_{\varepsilon_i \geq s}.
\end{aligned}$$

Note that this estimate holds for all ℓ , with all expressions vanishing when $\ell < p(n-p)$. Using the previous estimate and using (5.0.1), one finds the estimate

$$\begin{aligned}
0 &\leq \frac{1}{(2x)^{p(q-p)/2}} E_{x,p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) \\
&\quad - e^{-px} \sum_{\ell \geq 0} \frac{(px)^\ell}{\ell!} E^{\ell,p} \left(\prod_{1 \leq i \leq p} (\varepsilon_i - s)^{q-p} I_{\varepsilon_i \geq s} \right) \\
&\leq \sum_{\substack{0 \leq k_i \leq q-p \\ 1 \leq i \leq p \\ \sum k_i \geq 1}} \left(\frac{q}{\sqrt{2x}} \right)^{\sum k_i} \prod_{i=1}^p \binom{q-p}{k_i} \\
&\quad \times E_{x,p} \left(\prod_{i=1}^p \left(\frac{\lambda_i - x}{\sqrt{2x}} - s \right)^{q-p-k_i} I_{(\lambda_i - x)/\sqrt{2x} \geq s} \right) \\
&= O\left(\frac{1}{\sqrt{2x}}\right) \quad \text{for } x \rightarrow \infty
\end{aligned}$$

because the expectation $E_{x,p}$ tends to an integral, by Johansson's Proposition 5.1, applied to the function

$$g(x_1, \dots, x_p) = \prod_{i=1}^p (x_i - s)^{q-p-k_i} I_{[s, \infty)}(x_i).$$

For $0 \leq \ell < p(n-p)$, the expectation is automatically zero. Then

$$\begin{aligned}
&\lim_{x \rightarrow \infty} \frac{1}{(2x)^{p(q-p)/2}} E_{x,p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) \\
&= \lim_{x \rightarrow \infty} e^{-px} \sum_{\ell \geq 0} \frac{(px)^\ell}{\ell!} E^{\ell,p} \left(\prod_{1 \leq i \leq p} (\varepsilon_i - s)^{q-p} I_{\varepsilon_i \geq s} \right) \\
&= \lim_{x \rightarrow \infty} e^{-px} \sum_{\ell \geq 0} \frac{(px)^\ell}{\ell!} E^{\ell,p} \left(\prod_{1 \leq i \leq p} \left(\frac{\lambda_i - x}{\sqrt{2x}} - s \right)^{q-p} I_{(\lambda_i - x)/\sqrt{2x} \geq s} \right) \\
&= \lim_{x \rightarrow \infty} E_{x,p} \left(\prod_{1 \leq i \leq p} \left(\frac{\lambda_i - x}{\sqrt{2x}} - s \right)^{q-p} I_{(\lambda_i - x)/\sqrt{2x} \geq s} \right) \\
&= p! \int_{x_1 > \dots > x_p} \prod_{1 \leq i \leq p} ((x_i - s)^{q-p} I_{[s, \infty)}(x_i)) \frac{1}{Z_p} \Delta_p(x)^2 \prod_{j=1}^p e^{-x_j^2} dx_j,
\end{aligned}$$

by applying Proposition 5.1 to the continuous function

$$g(x_1, \dots, x_p) = \prod_1^p (x_i - s)^{q-p} I_{[s, \infty)}(x_i).$$

This leads to the ratio of Hermitian matrix integrals in (1.0.18). The numerator is precisely the matrix integral (3.0.1), which according to Theorem 3.1 satisfies the Painlevé equation (3.0.2), which ends the proof of Theorem 1.4. \square

6 Painlevé IV as a limit of Painlevé V

According to Theorem 1.4, the limit (1.0.18) leads to a solution of Painlevé IV; this was established by identifying the limit as a matrix integral and then applying Theorem 3.1. In this section we give an alternative proof of this fact, by directly taking the scaling limit of the Painlevé V equation (1.0.14). Remember from Proposition 1.1 and Theorem 1.4, we have

$$(6.0.1) \quad \prod_{i=1}^p \frac{(n-i)!}{(q-i)!} x^{-(n-p)p} e^{px} E_{x,p} \left(I_{\{\lambda \supseteq \mu\}}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) \\ = \exp \int_0^x \frac{u(y) - p(n-p) + py}{y} dy$$

and, with the rescaling

$$(6.0.2) \quad n - p = x + \sqrt{2xs},$$

we have from section 5, that

$$(6.0.3) \quad \lim_{x \rightarrow \infty} \frac{1}{(2x)^{p(q-p)/2}} E_{x,p} \left(I_{\lambda \supseteq \mu}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{ij}^\lambda \right) = c \exp \int_0^s h(y) dy,$$

for an appropriate choice of $h(y)$ and c . Taking into account the inverse of the rescaling (6.0.2), which for large n reads

$$(6.0.4) \quad x = n - s\sqrt{2n} + o(1)$$

and taking the logarithmic derivatives of both equations (6.0.1) and (6.0.3), we have

$$\frac{\partial}{\partial s} \log E_{x,p} \left(I_{\{\lambda \supseteq \mu\}}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) = \frac{u(x)}{x} \frac{\partial x}{\partial s}$$

$$\frac{p(q-p)}{2x} \frac{\partial x}{\partial s} + \frac{\partial}{\partial s} \log E_{x,p} \left(I_{\{\lambda \supseteq \mu\}}(\lambda) \prod_{\substack{(i,j) \in \lambda \\ n-q < j \leq n-p}} h_{(i,j)}^\lambda \right) \simeq h(s),$$

for large, but fixed n .

Letting $n \rightarrow \infty$ in both equations and keeping the leading terms lead to

$$\frac{\partial}{\partial s} \log E_{x,p} \simeq -\sqrt{\frac{2}{n}} u(n - s\sqrt{2n})$$

$$\frac{\partial}{\partial s} \log E_{x,p} \simeq h(s).$$

So, setting

$$h(s) := -\sqrt{\frac{2}{n}} u(n - s\sqrt{2n}),$$

we have

$$u(x) = -\sqrt{\frac{n}{2}} h\left(\frac{n-x}{\sqrt{2n}}\right)$$

and thus

$$u'(x) = \frac{1}{2} h', \quad u''(x) = -\frac{1}{2\sqrt{2n}} h'', \quad u'''(x) = -\frac{1}{4n} h'''.$$

Setting the rescaling (6.0.4) in the polynomials $Q(x)$ and $R(x)$, as defined in (1.0.15), we have

$$4Q = n(-2s^2 + 4(2p - q)) + o(n)$$

$$4Q' = 2\sqrt{2ns} + o(\sqrt{n})$$

$$2R = 2np(p - q) + o(n).$$

Substituting into (1.0.14), keeping the leading terms, which are of order n , and multiplying by 4, one finds the equation (1.0.19),

$$(6.0.5) \quad h''' + 6h'^2 - 4(y^2 + 2(q - 2p))h' + 4yh - 8(q - p)p = 0.$$

7 Appendix: Chazy classes

In his classification of differential equations

$$f''' = F(z, f, f', f''), \quad \text{where } F \text{ is rational in } f, f', f''$$

and locally analytic in z ,

subjected to the requirement that the general solution be free of movable branch points, Chazy found thirteen cases, the first of which is given by

$$(7.0.1) \quad f''' + \frac{P'}{P}f'' + \frac{6}{P}f'^2 - \frac{4P'}{P^2}ff' + \frac{P''}{P^2}f^2 + \frac{4Q}{P^2}f' - \frac{2Q'}{P^2}f + \frac{2R}{P^2} = 0$$

with arbitrary polynomials $P(z)$, $Q(z)$, $R(z)$ of maximal degree 3, 2, 1 respectively. Cosgrove and Scoufis [10, 9], (A.3), show that this third order equation has a first integral, which is second order in f and quadratic in f'' ,

$$(7.0.2) \quad f''^2 + \frac{4}{P^2} \left((Pf'^2 + Qf' + R)f' - (P'f'^2 + Q'f' + R')f \right. \\ \left. + \frac{1}{2}(P''f' + Q'')f^2 - \frac{1}{6}P'''f^3 + \delta \right) = 0;$$

δ is the integration constant. The relevant examples in this paper are:

P	$4Q$	$2R$	Painlevé eqt
1	$-4(u^2 + 2(a - n))$	$-8an$	P IV
u	$-u^2 + 2u(2n + a - b) - (a + b)^2$	$-bn(a + b + u)$	P V

Equations of the general form

$$f''^2 = G(x, f, f')$$

are invariant under the map

$$x \mapsto \frac{a_1 z + a_2}{a_3 z + a_4} \quad \text{and} \quad f \mapsto \frac{a_5 f + a_6 z + a_7}{a_3 z + a_4}.$$

Using this map, the polynomial $P(z)$ can be normalized to

$$P(z) = z(z - 1), z, \text{ or } 1.$$

In this way, Cosgrove shows (7.0.2) is a master Painlevé equation, containing the 6 Painlevé equations. In the cases of PIV and PV, the canonical equations are respectively:

$$g''^2 = -4g'^3 + 4(zg' - g)^2 + A_1g' + A_2 \quad (\text{Painlevé IV})$$

$$(zg'')^2 = (zg' - g)(-4g'^2 + A_1(zg' - g) + A_2) + A_3g' + A_4. \quad (\text{Painlevé V})$$

References

- [1] Adler M. and van Moerbeke P.: Hermitian, symmetric and symplectic random ensembles: PDE's for the distribution of the spectrum. *Ann. of Math.* **153** (2001), 149–189 (with M. Adler) (arXiv: math-ph/0009001)
- [2] Adler M. and van Moerbeke P.: Integrals over classical groups, random permutations, Toda and Toeplitz lattices. *Comm. Pure Appl. Math.* **54** (2000), 153–205 (arXiv: math.CO/9912143)
- [3] Adler M. and van Moerbeke P.: Integrals over Grassmannians and Random permutations. *Adv. in Math.* **181** (2003), 190–249 (arXiv: math.CO/0110281)
- [4] Aldous D. and Diaconis P.: Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem. *Bull. Amer. Math. Soc. (new series)* **36** (1999), 413–432
- [5] Aomoto K.: Jacobi polynomials associated with Selberg integrals. *SIAM J. Math. Anal.* **18** (1987), 545–549
- [6] Borodin A., Olshanski G.: Z -Measures on partitions, Robinson-Schensted-Knuth correspondence, and $\beta = 2$ random matrix ensembles. “Random matrices and their applications”, Mathematical Sciences Research Institute Publications #40. Cambridge University Press, 2001 (MathCO/9905189)
- [7] Borodin A., Olshanski G.: Distributions on partitions, point processes, and the hypergeometric kernel. *Comm. Math. Phys.* **211** (2001), 335–358 (math.RT/9904010)
- [8] Borodin A., Okounkov A., Olshanski G.: Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc. (to appear)* (math.CO/9905032)
- [9] Cosgrove C. M.: Chazy classes IX–XII of third-order differential equations. *Stud. Appl. Math.* **104** (2000), 171–228
- [10] Cosgrove C. M., Scoufis G.: Painlevé classification of a class of differential equations of the second order and second degree. *Stud. Appl. Math.* **88** (1993), 25–87
- [11] Dueñez E.: Random matrix ensembles associated to compact symmetric spaces. Thesis, Princeton University 2001
- [12] Forrester P. J. and Witte N.: Applications of the τ -functions of Painlevé equations to random matrices: P V, P III and the LUE, JUE and CUE. *Comm. Pure Appl. Math.* **LV** (2002), 679–727
- [13] Forrester P. J. and Witte N.: Application of the τ -function theory of Painlevé equations to random matrices: P VI, the JUE, CyUE, cJUE and scaled limits. *Nagoya Math. J.* **174** (2004), 29–114
- [14] Greene C.: An extension of Schensted's theorem. *Adv. in Math.* **14** (1974), 254–265
- [15] Johansson K.: Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)* **153** (2001), 259–296
- [16] Kaneko J.: Selberg integrals and hypergeometric functions associated with Jack polynomials. *SIAM J. Math. Anal.* **24** (1993), 1086–1110
- [17] MacDonald I. G.: Symmetric functions and Hall polynomials. Clarendon Press, 1995

- [18] Muirhead R. J.: Aspects of Multivariate Statistical Theory. Wiley series in probability and mathematical statistics, 1982
- [19] Sagan B. E.: The Symmetry Group. Wadsworth & Brooks, Pacific Grove, California 1991
- [20] Stanley R. P.: Enumerative Combinatorics. Cambridge Studies in Advanced Math. 49, Vols 1 and 2, 1997
- [21] Stanley R. P.: Some combinatorial properties of Jack symmetric functions. Adv. in Math. **77** (1989), 76–115
- [22] Tracy C. A. and Widom H.: On the distribution of the lengths of the longest monotone subsequences in random words. Probab. Theory Related Fields **119** (2001), 350–380
- [23] van Moerbeke P.: Integrable lattices: random matrices and random permutations. “Random matrices and their applications”. Mathematical Sciences Research Institute Publications #40. Cambridge University Press, 2001, pp 321–406 (<http://www.msri.org/publications/books/Book40/files/moerbeke.pdf>)

Received January 3, 2006

Department of Mathematics, Brandeis University, Waltham, Mass 02454, USA
adler@brandeis.edu

Department of Mathematics, California Institute of Technology, Pasadena, CA 91125, USA
and Clay Mathematics Institute Long-term Prize Fellow
borodin@caltech.edu

Department of Mathematics, Université de Louvain, 1348 Louvain-la-Neuve, Belgium and
Brandeis University, Waltham, Mass 02454, USA
vanmoerbeke@math.ucl.ac.be and @brandeis.edu